

# Alignment Analysis of Michigan Grades 3-8 English Language Arts Standards and Assessments

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Alignment describes the match between expectations and assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Language Arts and Science Education* (Webb, 1997).

Dr. Norman Webb from the University of Wisconsin led an alignment institute in Lansing, Michigan, September 21st through 23rd, 2005. The institute included twelve reviewers, including language arts content experts, district language arts supervisors, and language arts teachers. Eight reviewers were from Michigan, and four were experts with experience from other states. Over the three days, the reviewers analyzed the agreement between the state's language arts standards for grades 2 through 7 and the fall 2005 assessments for grades 3 through 8.

## Summary

The alignment between the assessments and the language arts standards at each grade was acceptable. The over-emphasis on one or two reading objectives on the assessment is not a critical alignment issue, since all of the other alignment criteria were fully met. The alignment between the assessments and the writing standard at each grade needs slight improvement. One to three more objectives for each grade need to have at least one corresponding item for the assessments to fully meet the Range-of-Knowledge Correspondence criterion. Reviewers were very consistent in assigning items to standards, but showed less consistency in assigning items to specific grade-level expectations. This implies some overlap in content covered by the grade-level expectations, or lack of clarity in the written statements. Because reviewers found it difficult to distinguish among many of the objectives, this lowered the reviewer agreement on the precise objective measured by an item. The reviewers observed that the coverage of content on the assessments improved over the grades. Reviewers indicated there were some very challenging items on the grade 7 and grade 8 assessments.

## Alignment Criteria

For the purposes of this analysis, we have employed the convention of standards, goals, and objectives to describe three levels of expectations for what students are to know and do. Standard is used here as the most general (for instance, *Reading*) of the terms. There are two such standards for grades 2 through 7, *Reading* and *Writing*, that are assessed at the state level. Each standard is comprised of four to six goals, each of which is comprised of up to eight grade-level expectations (objectives). These objectives are intended to span the content of the goals and standards under which they fall.

Reviewers were trained to identify the depth-of-knowledge (DOK) of objectives and assessment items. This training included reviewing the definitions of the four DOK levels and reviewing examples of each. Then the reviewers participated in 1) a consensus process to determine the DOK levels of the Michigan content objectives and 2) individual analyses of the assessment items of each of the six assessments.

To derive the results on the degree of agreement between the Michigan language arts standards and each assessment, the reviewers' responses are averaged. Any variance among reviewers is considered legitimate, with the true DOK level for the item falling somewhere between two or more assigned values. Such variation could signify a lack of clarity in how the objectives were written, the robustness of an item that can legitimately correspond to more than one objective, and/or a depth of knowledge that falls in between two of the four defined levels. Reviewers were allowed to identify one assessment item as corresponding to up to three objectives—one primary hit (objective) and up to two secondary hits. However, reviewers could only code one DOK level to each assessment item, even if the item corresponded to more than one objective. Finally, in addition to learning the process, reviewers were also asked to provide suggestions for improving it.

Reviewers were instructed to focus primarily on the alignment between the state standards and the various assessments. However, they were encouraged to offer their opinions on the quality of the standards, or of the assessment activities/items, by writing a note about the item. Reviewers could also indicate whether there was a source-of-challenge issue with the item—i.e., a problem with the item that might cause the student who knows the material to give a wrong answer, or enable someone who does not have the knowledge being tested to answer the item correctly. For example, a language arts item that requires specialized knowledge from another area, such as science or mathematics, may represent a source-of-challenge issue because the skill required to answer is more than a reading comprehension skill.

The results produced from the institute pertain only to the issue of agreement between the Michigan state standards and the six grade-level instruments for fall 2005. Note that this alignment analysis does not serve as external verification of the general quality of the state's standards or assessments. Rather, only the degree of alignment is discussed in these results. The averages of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the

averages lessened the error that might result from any one reviewer’s finding. Standard deviations are reported, which give one indication of the variance among reviewers.

To report on the results of an alignment study of Michigan’s curriculum standards and six different standardized assessments, the study addressed specific criteria related to the content agreement between the state standards and grade-level assessments. Four alignment criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation. These are defined briefly in Table 1.

Table 1  
Criteria for Alignment

<b>Criterion</b>	<b>Definition</b>
Categorical Concurrence	<i>At least six items measuring content from a standard</i>
Depth-of-Knowledge Consistency	At least 50% of the items corresponding to a standard had to be at or above the level of knowledge of the standard
Range-of-Knowledge Correspondence	Fifty percent of the benchmarks for a standard had to have at least one related assessment item
Balance of Representation	Items/activities are distributed among all of the benchmarks at least to some degree

## **Alignment of Curriculum Standards and Assessments**

The results from the alignment analysis for the six grades are presented in Table 2. “Yes” indicates that an acceptable level on the criterion was fully met. “WEAK” indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. “NO” indicates that the criterion was not met by a noticeable margin.

The grade 3 language arts assessment had 37 items. The assessments for the other five grades each had 45 items. The distribution of items was consistent across grades, with about 80% of the items measuring content related to reading and 20% of the items measuring content related to writing.

The alignment between the language arts assessments and the previous grade’s standards was acceptable for reading and, generally, good for writing. For both of these content areas, three of the four alignment criteria were fully met (Table 2). The alignment between the reading assessments and standards had an acceptable level for Categorical Concurrence, Depth-of-Knowledge Consistency, and Range-of-Knowledge Correspondence. This means that for each grade level the assessment had an adequate number of items at a sufficient level of complexity and that covered an ample proportion of the objectives.

Table 2

*Summary of Acceptable Levels on the Four Alignment Criteria for Language Arts Grades 3–8 Assessments for Michigan Alignment Analysis*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
Grade 3				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	NO	WEAK
Grade 4				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	WEAK	YES
Grade 5				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	NO	YES
Grade 6				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	WEAK	YES
Grade 7				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	WEAK	YES
Grade 8				
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	WEAK	YES

## Action Needed for Assessments and Standards to be Fully Aligned

In summary, because the alignment between the assessments for grades 3–8 and the standards for grades 2–7 is reasonable, only slight changes in the assessment of writing are needed to achieve full alignment. To achieve full alignment would require these changes or modifications:

Grade 3. Three items need to be replaced by items that measure grade-level expectations for writing that are not currently assessed. Five of the eight items coded by reviewers as assessing writing corresponded to one grade-level expectation, W.GR.1 (grammar and usage). The reviewers did not consistently agree on what grade-level expectations the other three items measured, but they did agree that these items measured content knowledge related to the writing process. Full alignment could be achieved by replacing three of the items related to grammar and usage with items that more clearly

measured grade-level expectations under the writing process, writing genres, and spelling. This action would also improve the balance.

Grade 4. One writing item needs to be replaced by an item that clearly relates to a writing objective not currently measured. For example, one of items 53, 54, or 56 that currently assess content knowledge related to grammar could be replaced by an item measuring content related to Objective W.PR.6.

Grade 5. Two items need to be replaced by items that clearly relate to a writing objective not currently measured. Reviewers coded fewer items as corresponding to writing objectives at grade 5 than other grades. Therefore, it would be better for two reading items to be replaced by items that measure writing objectives not currently assessed.

Grades 6, 7, and 8. One writing item needs to be replaced by an item that clearly relates to a writing objective not currently measured.